DOCUMENT RESUME

ED 134 540 95 SP 010 703

AUTHOR Pottinger, Paul; Klemp, George
TITLE The Fund for the Improvement of Postsecondary Education. Final Report.
INSTITUTION McBer and Co., Boston Mass.
SPONS AGENCY Fund for the Improvement of Postsecondary Education (DHEW), Washington, D.C.
PUB DATE Oct 75
CONTRACT OEC-0-74-9258
NOTE 39p.
AVAILABLE FROM Institute for Competence Assessment, McBer & Company, 137 Newbury Street, Boston, Massachusetts 02116 ($1.50)

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Academic Achievement; Cognitive Measurement; Educational Assessment; Evaluation Criteria; Evaluation Methods; Evaluation Needs; Higher Education; Measurement Instruments; *Measurement Techniques; *Performance Based Education; *Post Secondary Education; *Relevance (Education); Research Needs; *Skills; Test Construction; Test Validity

ABSTRACT

This document reports efforts to address some of the problems of competency-based assessment by developing and utilizing unique measures of competence. Seven conceptual guidelines are identified for test development and utilization: (1) new competencies must be identified and operationally defined; (2) new competencies should have general significance to a wide variety of career and life outcomes; (3) new definitions of competencies and measures developed for their assessment should be easy for faculty and students to understand; (4) competencies should be empirically linked to external realities; (5) the discovery of new ways of measuring competencies should acknowledge levels of performance required for entry into roles outside of academic settings; (6) standards for awarding credentials should acknowledge levels of performance required for entry into roles outside of academic settings; and (7) new attempts to define and assess learning outcomes should not be guided solely by attempts to make them functionally equivalent substitutes for traditionally assessed school achievement. A number of measures were developed along these guidelines; they are organized according to three outcome domains--cognitive, effective, and social. Descriptions of these measures are presented, and their applicability for competency based education and use in postsecondary institutions is discussed. Finally, the critical concept of the meaning of measurement is discussed along with some additional problems to be faced by educators, researchers, and funding agencies if the problems of assessment in higher education are to be adequately redressed. (MM)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED134540

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Paul Pottinger

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

FINAL REPORT

to

The Fund for the Improvement
of Postsecondary Education

Contract # OEC-O-74-9258
October 1975

Paul Pottinger, Ph.D.
George Klemp, Ph.D.

Institute for Competence Assessment
Division of McBer & Company
137 Newbury Street
Boston, Massachusetts 02116
617 261 5570

SP010 123

© McBer & Company, 1975

ICA

TABLE OF CONTENTS

		Page
I.	Introduction	1
II.	Guidelines for Test Development and Utilization	3
III.	Competency-based Measures	7
	Measures of Cognitive Outcomes	8
	Measures of Effective Outcomes	13
	Figure 1: A General Integrative Model	14
	Measures of Social Outcomes	16
	Table 1: Relationship of Eighteen Competency-based Measures to Three General Competency Domains	19
IV.	The Meaning of Assessment Measures	21
	Epilogue: What's in a Name?	25
	Bibliography	27
	Appendix A	29
	Appendix B	31
	Appendix C	35

I. INTRODUCTION

Competency-based education (CBE) is a relatively new approach to answering the challenge faced, but not met, by traditional education: to teach those skills that help one to be successful in life and in one's life work rather than merely to be successful in an academic setting. The unique quality of competency-based education is that it teaches and measures "competencies" rather than basing education and assessment on courses taken, time invested, or credits earned. The critical element of competency-based education is that real life competencies can be defined so that:

- goals are clearly articulated;
- outcomes can be accurately communicated and measured;
- students know what is expected of them;
- tests are valid and reliable and can be used to give concrete feedback to students about how well they are doing; and
- instructors are confident that what is taught, the growth of the student, and the measures to assess the growth are all relevant to the ability to do, i.e., the ability to function adequately, appropriately, and confidently in life.

As competency-based education and other innovative mechanisms are used for awarding credentials, there is an increased need for reliable, valid, and cost-effective measures. These new measures must be responsive to both traditional and newly defined learning outcomes, which are related to success outside of the world of academia. Standard achievement, knowledge, ability and aptitude tests have proven inadequate in measuring the skills, abilities, and characteristics that are predictive of success outside of the classroom, whether such real competencies are attained in institutions of higher education or from other life experiences. In other words, standard methods of educational evaluation measure a very limited and specialized type of competence that is unrelated to important life outcomes such as occupational success or life adjustment. Educational innovations are significantly affecting competencies that simply cannot be measured in traditional ways or with traditional tests. Thus, both traditional and non-traditional educational programs, which are designed to better prepare people for work and life, are in need of measures of competencies germane to success in life outside of academia. Furthermore, since the competency-based approach to education makes the demonstration of competence the sine qua non for the award of credentials,

the measurement of attained competence has become the single most important problem in the effective implementation of these programs.

Response to these problems of assessment procedures unique to competency-based education have not kept pace with the proliferation of experimental programs to implement new approaches to teaching and learning.

McBer and Company, under a contract with the Fund for The Improvement of Postsecondary Education, attempted to address some of the problems of competency-based assessment by developing and utilizing unique measures of competence. The issues of competency-based assessment were seen as important and distinctive from other research done by McBer. This resulted in the creation of a separate division of McBer and Company known as the Institute for Competence Assessment (ICA). The ICA division of McBer is dedicated solely to improving the state of the art of competence assessment in higher education, business, and other public and private sectors.

The remainder of this report deals with the conceptual guidelines used by McBer's Institute for Competence Assessment in developing or utilizing these measures (II), presents descriptions of the measures and areas of their applicability for CBE and other postsecondary institutions (III), discusses the critical concept of the meaning of measurement and addresses some additional problems that must be faced by educators, researchers, and funding agencies if the problems of assessment in higher education are to be adequately redressed (IV).

II. GUIDELINES FOR TEST DEVELOPMENT AND UTILIZATION

These guidelines are partially based upon three aspects of assessment which have been addressed elsewhere by Pottinger (1975):

- The identification and definition of competencies relevant to life and work outside of academia;
- Instrumentation, techniques, and processes of evaluation that provide reliable and valid measures of these competencies; and
- Standardization, and/or establishment of levels of performance necessary and sufficient for awarding credentials.

A. "New" competencies must be identified and operationally defined.

There are many outcomes of the learning experience that have greater validity than grades in school as a basis for awarding credentials. Those that have been identified have been accepted as being important and meaningful in establishing a person's competence, yet many academicians have not sought to operationalize, measure, and award credit for many of these learning outcomes. There are many other criteria than traditionally-rewarded scholastic achievement that are important as competencies in the practical world, and most of them are as yet unidentified.

B. New competencies should have general significance to a wide variety of career and life outcomes.

Competencies cannot be meaningfully defined by a seemingly endless reduction of specific skills, tasks and actions which ultimately fall short of real-world requirements for effective performance. In fact, the more essential characteristics for success often turn out to be broad or generalized abilities or characteristics which are sometimes more easily operationally defined and measured than an array of specific "subskills" which do not add up to general competence.

C. New definition of competencies and measures developed for their assessment should be easy for faculty and students to understand and use.

New competency definitions should be readily recognizable as important, and related assessment techniques and instruments should be easy for faculty and students to understand. It is necessary to guard against competency definitions and measures that are so complex, or trivial, or esoteric that students and faculty can neither understand them, nor accept them as meaningful and useful. In other words, educational goals should

not be rendered unintelligible, and assessment procedures and instruments should not mystify the process of evaluating student progress.

D. Competencies should be empirically linked to external realities.

Many educators assume that such things as the ability to master new bodies of knowledge quickly and effectively, to analyze and solve problems, to develop new skills efficiently, and to utilize knowledge are prerequisites for individuals if they are to take advantage of life's opportunities and surmount its difficulties. What is missing are the measures of these general abilities, which are related to important life outcomes. Only when we know what makes the difference between adequate and inadequate performance, based on empirical analyses of professions and other life activities, will we be able to develop or improve such measures, clarify new competencies, and establish credentials of demonstrated value.

E. The discovery of new ways of measuring abilities (competencies) is needed.

The measurement technology must be innovative and new, not just a new name for traditional procedures. Paper-and-pencil (objective) tests, due to method variance, correlate better with each other than they do with performance criteria. If post-secondary education is to break out of this closed circuit, different approaches to testing must be sought in areas such as learning, critical thinking, problem solving and other newly defined competencies. Measures of competence must require that the test taker generate appropriate learning outcome responses. The primary learning objectives of education is not to help an individual select from among a set of predetermined alternatives. Rather, it is to enable a person to know how to reason; how to marshal evidence for or against an hypothesis; how to analyze a problem into its components; how to see similarities and differences in objects, ideas, and events; how to partial out crucial information from the trivial; and how to integrate these skills with purpose and meaning. Multiple-choice tests do not and cannot measure these abilities. And the behavioral observation/documentation approaches that are popular in experiential learning assessment do not allow these abilities to be measured with adequate reliability or validity. (For a brief critique of popular new approaches to measuring competencies within the competency-based education movement, see Section IV, pp. 21 and 22, and also Appendix A.)

F. Standards of performance for awarding credentials should acknowledge levels of performance required for entry into roles outside of academic settings.

The establishment of criteria or standards of competence is one of the most difficult problems to be addressed. In every case, where standards of competence are determined for new or for more traditional outcomes, appropriate levels should be established by sufficient empirical evidence to ensure that they will not be viewed as arbitrary. Many educators are satisfied with a priori judgments of what skills and levels of performance are adequate. It is startling to realize how much we accept the face validity of credentials and how little we really know about (1) the correspondence between the abilities and levels of performance that these credentials represent, and (2) what is needed for adequate performance in life's tasks. We need to develop better benchmarks for evaluating the standards and the offerings of postsecondary institutions. (For further comments about the establishment of standards and levels of performance, see Appendix B.)

G. New attempts to define and assess learning outcomes should not be guided solely by attempts to make them functionally-equivalant substitutes for traditionally assessed school achievement.

Competency-based education requires a different type of evaluation from traditional programs to the extent that learning outcomes differ in significant ways. For example, learning outcomes in CBE are often defined in terms of what a person can do, not merely in terms of what one knows. Furthermore,

> "whereas in traditional programs evaluation is primarily linked to the credentialing process, in competency programs it is also used as a formative teaching tool. In other words, students are made aware of the criteria and standards for certification in a competency, and their progress is frequently measured so that help can be provided as necessary. Assessment that simply places students in a percentile or just discriminates between passing and failing is not adequate for competency-based programs. Formative diagnostic advice is needed--information that tells if the student is 'real world' competent" (Hodgkinson, 1975).

The temptation to restrict the development of new measurement instruments, techniques, and procedures, in order to achieve comparability with those that have gone before, has

great political appeal for making such innovations palatable to traditionalists. However, if institutional and credential reforms are to succeed, we need to move beyond the recognized limitations of traditional assessment systems.

The implications of these guidelines for research and change are numerous (See Appendix C). Postsecondary educators are in strong agreement that certain abilities and characteristics are necessary for success in life. Traditional and non-traditional curricula have been focused on these generic abilities, but few people have empirically validated measures of these characteristics. Educators often accept, on faith, certain abilities as critical to successful performance in life--such as the ability to learn new information efficiently, to utilize knowledge, to observe, to analyze and solve problems, to be pro-active rather than merely reactive, to be empathic, and to integrate all of the above skills. What is needed are measures of these general characteristics which are causally linked to important life outcomes.

The efforts involved in this project have constituted a response to the need for new measures appropriate to the learning outcomes for competency-based education programs. These learning outcomes do not differentiate general education from more career-oriented programs of learning. Rather, they differentiate programs whose learning objectives and methods and standards for evaluation are clearly specified from those programs whose learning goals, although rational, are vague with respect to how criteria and standards for excellence are determined and evaluated.

III. COMPETENCY-BASED MEASURES

Psychologists have often failed to develop measuring instruments that are sensitive enough to detect effects of primary interest to educators (see above). According to McClelland (1976) there is ample reason to believe that educational psychologists have unnecessarily restricted the range of methods they have employed to measure the impact of higher education. Time-saving and money-saving incentives have resulted in a predominance of measures which utilize the multiple-choice questionnaire format. The consequences of this decision, according to McClelland, have been far greater and more limiting than most people realize.

It would serve us well to ask the extent to which a multiple choice or a true-false test has any bearing on what people do in real life and on the competencies that they possess. In our daily lives we are constantly called upon to process various kinds of information, to analyze its components, to associate this new information with that which we have stored away in our memory, to partial out the crucial information from the trivial, and to integrate this information into our cognitive structure. In this way, we constantly use information from many sources to solve problems, and in the process we learn new things about our world and ourselves. In truth, people are almost never asked to recognize a correct answer among a list of three or four alternatives. Rather than being reactive to such well-defined situations, people must be pro-active in situations which provide only partial information.

The one thing most traditional testing methods have in common, regardless of what they purport to assess, is this: they only measure one's ability to retrieve information after it has been stored. And many such methods fail even in this; a multiple-choice test, for example, measures the ability to recognize rather than recall. Essay tests are very subjectively scored, even when there is only one "correct" answer or line of reasoning as is often the case.

Storage and retrieval of information are not the important issues for higher education. Indeed, Ebbinghaus demonstrated many years ago that 70 percent of that which is learned in the classroom is forgotten within one year. Rather, the issue is a more substantive one; how is the knowledge gained in coursework used to come to grips with the practical problems of living. Implicit in this are three related issues of particular importance: how able are people in <u>processing</u> new information for problem solving; how able are they in <u>integrating</u> this information to form new solutions; and how able are they in <u>implementing</u> these solutions.

A number of measures have been developed by McBer to answer the need for a more "pro-active" measurement technology to assess the factors of process, integration, and implementation. For the sake of clarity, and consistent with the competency-based orientation toward outcome-relatedness, the measures described below are organized according to three outcome domains: cognitive, effective and social outcomes.

Cognitive outcomes. Measures in this domain assess characteristics purportedly measured by traditional tests of mental ability, aptitude, and knowledge. The differentiating characteristic between McBer measures and traditional tests is that McBer measures are based on the idea that the test taker should provide all the information necessary for adequate and appropriate response to a problem on a test, as opposed to merely selecting from a set of prepared alternative responses.

Effective outcomes. Variables measured in this domain are directly translatable to behavior patterns required beyond the world of academia. This category is derived from White's (1961) term "effectance," which means positive, goal-directed and productive interaction with and influence on the environment.

Social outcomes. These measures assess areas of interpersonal competence which often facilitate the fruition of cognitive and effective dimensions of competence in life. They take into consideration the attitudes, values, and orientations toward others which moderate life goals and the means for achieving them.

Measures of Cognitive Outcomes

1. Test of Critical Thinking. The ability to analyze new information and to synthesize new concepts based on this information reflects the ability to integrate information into one's own cognitive structure. As the cognitive structure grows, so does the ability to think critically, to make a cogent argument, and to reason inductively; thus, the Test of Critical Thinking is a measure of cognitive development. The test takes the form of two sets of stories which an individual is asked to compare thematically. This "thematic analysis" is scored according to nine categories of critical thinking and a total score is derived. This scoring system is reliable, efficient and cost-effective. Each scoring category is a logical and independent dimension of critical thinking skill.

This test, developed by Winter (1973), is distinguished from other measures of critical thinking skills in that it demands the test-taker to actually produce critical arguments,

rather than to simply recognize the critical elements of arguments presented to him. This instrument can be used to chart a student's progress in learning this skill. Alternative versions of the test have been developed to assess both the quality and structure of critical thinking.

2. Analysis of Argument. Given the ability to think in a critical fashion, a higher order skill is the ability to analyze other information which may or may not exemplify critical thought. This test requires proactive responses in the form of writing both a defense and a refutation of arguments which are based on what may be false assumptions, insufficient information, or unsubstantiated generalizations. The point of asking students to argue on both sides of the issue is to determine whether they can present an organized case, regardless of their feelings on the matter. This test is reliably scored according to a coding system, developed by Stewart (1976), which gives positive points for presenting an organized, logical case and negative points for simple enthusiastic endorsement or just stringing together unrelated facts that do not seem germane to the point of argument.

3. Concept Formation. The McBer Concept Formation Test is a programmed learning approach which is used to study the ability of people to learn concepts by comparing similarities and differences among objects. Concept formation is an important part of being able to incorporate new information into existing memory structures, to assimilate this information in such a way as to classify it in terms of the most important distinctive features. The ability to recognize elements of similarity and to identify a concept according to these elements is important, for example, in diagnosing a problem which shares the history of a difficult situation of the past, and thus being able to effectively ward off future trouble.

The Concept Formation test begins with a series of objects paired with a series of names which stand for the concepts to be learned. While the objects change over trials, objects representing the same concept have certain things in common (e.g., shape, generic class, numerousness). The speed and the accuracy with which the concepts are learned, adjusted for speed and accuracy of paired associate learning, yields a measure of concept formation.

4. Speed of Learning. The Speed of Learning test is an approach to measuring one's ability to process new information in a short span of time. The importance of this skill is self-evident. Adjustments to new situations, such as a change in course of study, job, or economic condition, must be made swiftly by an individual so that he can be effectively proactive in life. The Speed of Learning test is designed to

assess not only the ability to acquire general knowledge, but to acquire it selectively--that is, to remember the functionally important pieces of information rather than dwelling on insignificant pieces.

The form of the test is a presentation of the new material on a recorded tape, followed by a series of questions about what has been heard. Using an audio presentation allows material to be presented to each respondent at a defined rate, thereby ensuring that such factors as reading ability are not the real skills being tested. The questions, too, require the person to recall information, rather than to recognize the appropriate response from a set of alternatives. (Again, the real world does not supply a definite set of possible responses, one of which we know is correct). After the first set of questions, the material is presented a second time. This allows one to assess the effects on learning of repeated exposure to the same material, as well as providing an index of learning potential in a recall-type format. This second phase of the test provides a built-in validity check on the first phase, while it allows the assessor to chart the ability to learn new materials in three general areas--natural science, social science, and humanities--as well as in general biographical and process-oriented knowledge.

5. Learning Styles. A successful worker is distinguished not so much by any single set of knowledge or skills, but by the ability to adapt to and master the changing demands of one's job and career: that is, his ability to learn. Continuing success in a changing world requires an ability to explore new opportunities and learn from past successes and failures. Kolb's Learning Styles Inventory (1971) is a measure of individual learning styles which affect decision-making and problem-solving. The four styles, Concrete Experiential learning (CE), Reflective Observation (RO), Abstract Conceptualization learning (AC), and Active Experimentation learning (AE), when present in equal proportions, indicate the type of person who is able to involve himself fully, openly, and without bias in a new experience (CE), can reflect on and observe these experiences from many perspectives (RO), is able to create concepts that integrate his observations into logically sound "theories" (AC), and can use these theories to makes decisions and solve problems (AE) (Kolb, 1973).

6. Sayings Score. This procedure has the following format. First, a person is given a set of questions. Whether he knows the correct answers depends upon the individual's basic understanding of a principle or basic fact in the content area being tested. After answering these questions, he is given the answers, in the form of the principles or basic facts that define the correct responses. Finally, the individual is given a new set of questions; these questions are also derived from

the principles and facts in a way that is logical for anyone who is familiar with the content area being tested.

Traditional testing usually stops with the first set of questions. A competent person, however, may have an appropriate cognitive map or schema for a question, but may simply have forgotten the correct answer; or he may be familiar with a content area while the specific material is new to him. Nevertheless, once given the basic principles that underlie this first set of questions in a savings score test, the competent person has little difficulty in answering the second set of questions, since all the relevant cognitive schema have been activated by the first question-answer exercise. Thus, savings score tests incorporate both the processing and integrating functions of competency.

By contrast, for the naive or less competent person, the learning experience provided by the answers to the first question set is new, the information learned is easily forgotten, and no cognitive map is activated that makes answering the second set of questions possible. The untrained person cannot process the new information effectively, nor can he integrate it into an existing schema.

McBer has developed several prototypes of the savings score tests, including a test for general knowledge and a test for knowledge of human development.

7. Proactive Case Response. The purpose of this measure is to assess one's ability to use knowledge and cognitive skill for diagnosis, judgment and problem-solving. It serves to measure the ability to integrate information from one's existing knowledge base in response to a detailed situation or "case". Individuals are asked questions about the case which draw on their general knowledge of one situation. The people who are taking the test must (a) figure out what is happening, or diagnose the case, (b) decide what they should do to get a better idea of what is going on, and (c) respond in a way that demonstrates good judgment as to what should be done.

The test is not scored for "correct answers," since a case may have many valid interpretations, but rather for the appropriateness of a response. For example, if there are inconsistencies in the case or something is very wrong with the situation described, simply knowing that something is wrong and that certain action steps must be taken to find out what is wrong is as good as knowing precisely what is wrong in technical terms. Both kinds of responses are appropriate to the diagnosis of the situation and the implementation of recommendations. Accordingly, answers are coded based on an empirically-derived schema in which several responses are all scored as correct. The code is objective enough so that anyone

who learns it can score the test.

8. Programmed Cases. Based on incidents culled from in-depth interviews with criterion groups, programmed cases can be developed to test for social learning and judgment. Versions of this technique, developed for the U.S. Information Agency and the U.S. Navy, consist of a series of incidents to which several alternative responses are attached. All of the incidents pertain to a particular individual, or "case". "Distractors," or the incorrect responses, are developed with the aid of expert judges. The cases are programmed in such a way that a person with good judgement, i.e., who does not make snap, impulsive judgments, will become more accurate in his choices of the correct alternative as he proceeds through the case.

The programmed case technology has two primary uses:

- diagnostic assessment of how one uses information in making decisions about others or predicting their behaviors, and
- examination of the process by which decisions/predictions are made, including the analysis of values, biases, and preconceptions that interfere with veridical impressions of others and their situations.

9. General Integrative Model. Once an individual has gone through a series of academic or life experiences that enhance his competence in dealing with school, work, and other life situations, the task becomes that of measuring such generalized variables as the ability to cope with new problems, to find appropriate solutions, and to take correct action steps. The General Integrative Model requires an individual to demonstrate the following abilities:

- to observe;
- to extract relevant information;
- to analyze and integrate this information;
- to ask appropriate questions;
- to process new information in response to such questions;
- to utilize this information and one's knowledge in making sound and logical recommendations;
- to develop main and contingency plans;
- to set meaningful goals; and

- to feed back this new information into the process for better problem analysis and solutions.

This model is not a measure, per se, but a collection of measures logically ordered, to assess problem-solving skills. Figure 1 outlines a configuration of tests which represents one possible model of this approach. The tests themselves would be oriented toward a specific content area. The progress from stage to stage in the model presents the students with subproblems to solve, e.g., what new information to seek, what conclusions to draw, and what decisions to make derived from the information gathered at a given time.

Measures of Effective Outcomes

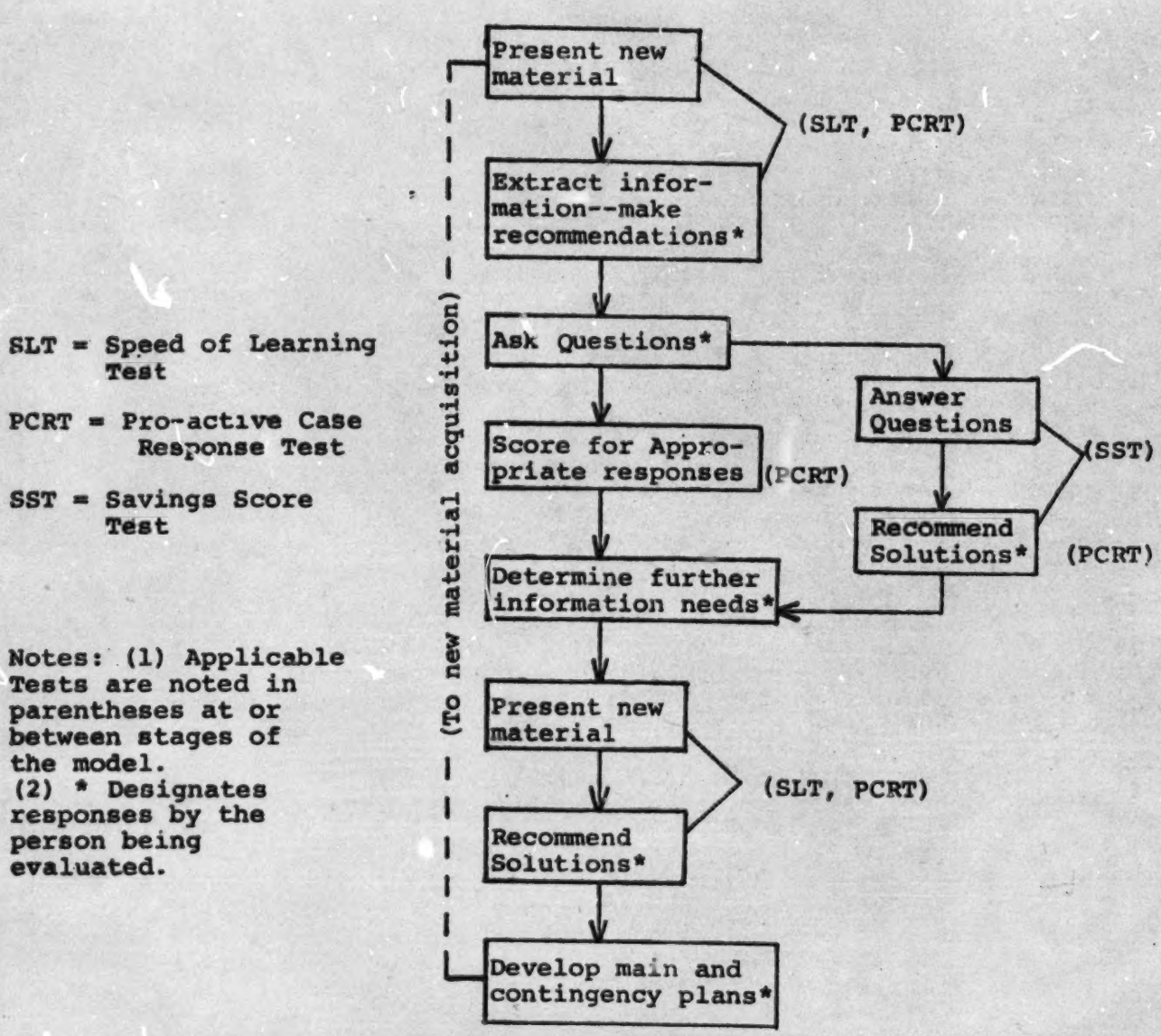
10. Diagnostic Listening. The Diagnostic Listening test consists of a taped presentation, with slides, of interviews with various individuals typical of the people one might encounter in social service work. People who take this test listen to an interview or a brief statement by a particular individual on the tape, and are then asked some questions about what has happened, what the person is really like, and what they would recommend for the person. This test requires the skills of listening, observing, and judging skills which have been found necessary in human service work.

There are two subscales in this test. The Casework Subscale, consisting of 42 items, is made up of four interviews, and after each of them the person taking the test is asked to answer questions and to make judgments on a multiple-choice answer sheet. The Positive Bias Subscale, consisting of 39 items, shows to test-takers three slides of clients of different sex and race with accompanying brief monologue. After each of these presentations, the test-takers are required to rate several adjectives as "does describe" or "does not describe" the client. An overall Positive Bias score is obtained by summing the number of positive yet realistic adjectives chosen. The Diagnostic Listening test measures faith in the client's ability to change, ability to observe and diagnose human problems, ability to set realistic goals, and ability to propose imaginative solutions.

General Comments related to tests 11-16. Much research has been accumulated by McClelland (1958, 1971), Atkinson (1958), and others that shows that thought patterns are related to important kinds of behaviors. The Exercise of Imagination is McBer's version of the Thematic Apperception Test (TAT) which is used to elicit thought patterns of the test-taker.

An individual taking the test is asked to write narratives to pictures. Each of these narratives addresses the following questions about the pictures: what is happening; who are the

FIGURE 1: A GENERAL INTEGRATIVE MODEL
(One approximation)



SLT = Speed of Learning Test

PCRT = Pro-active Case Response Test

SST = Savings Score Test

Notes: (1) Applicable Tests are noted in parentheses at or between stages of the model.
(2) * Designates responses by the person being evaluated.

people; what has happened in the past that has led to the situation; what is being thought--what is wanted by whom; what will happen; and what will be done. The stories are then scored, according to a prescribed set of codes or rules, to uncover certain patterns of thought that are expressed in the stories.

The link between thoughts and behavior has been repeatedly demonstrated to be strong, as opposed to the link between attitudes and behavior. The attitude-behavior link is influenced primarily by situational factors. An attitude may represent a specific goal or objective, but such goals and objectives may change according to situational demands and constraints. However, whether a specific goal changes or not, the characteristic style with which any goal is attained is determined to a large extent, by thought patterns which are relatively consistent within individuals.

The thought patterns scored in the following tests are particularly relevant to effective and social outcomes. Measures 11 through 16 are based on reliable scoring codes that can be applied to any written narrative which addresses the types of questions mentioned above.

11. Achievement Motivation. McClelland has shown in extensive research (1961) that people high in the need for achievement are practical and interested in efficiency--in short, they are good practical decision makers. They are independent, good at evaluating information for its practical utility, and original in the sense that they keep looking for better ways of doing things. For instance, they make good career decisions and regularly achieve greater success earlier in their careers. In a recent Harvard University longitudinal follow-up study (1976), freshman need-for achievement scores correlated with "early success" in various fields 14 years later.

12. Socialized Power. A major distinction in concern for power centers around whether a person is motivated to express or increase his own power, reputation, or glory without concern for others (personalized power), or whether he is drawn to seek power for the good of others or for the good of some cause (socialized power). For example, people high in socialized power are much more apt to be responsible citizens and to join voluntary organizations, often getting elected to office in them.

13. Stage IV Power. This power orientation, recently identified by McClelland (1975) is a concern for doing one's duty, that is, to be the instrument of a power which extends beyond the self.

14. Self-Definition/Cognitive Initiative. Self-definition/cognitive initiative is a general characteristic of an individual which encompasses the way one thinks about the world and himself, the way one reacts to new information, and the way one behaves. People with this competency are not only able to think clearly, but also to reason from the problem at hand to a solution, and to propose and take effective action on their own. Such competence is characteristic of people who think in a rational, systematic way on their own, and who can anticipate problems before they arise. In short, it might be said that people who are high in this characteristic are able on their own to see things clearly, to understand the causes of events, to reason from problem to solution, and to take effective action to solve problems. For example, the self-definition score has been quite useful in distinguishing between women who pursue careers following college and those who do not (Stewart, 1974.)

Measures of Social Outcomes

15. Affiliation Motive. While strong need for affiliation does not seem to be critical for effective task-oriented performance, and might actually be detrimental in some situations, recent research has suggested that some concern with the feelings of others, and with the compassionate quality of relationships, does seem to lead to superior capability in working with other people. Such basic affiliative concern is helpful in understanding others and in building good working relationships with colleagues and associates. This kind of affiliative concern is a means to attain other, broader kinds of satisfaction, and might well be labeled social sensitivity and skill.

16. Social-Emotional Maturity. Stewart's (1974) measure of social emotional maturity has been shown to be associated with managerial success and also with occupations which have a management component, e.g., Human Service Workers. This competency is also measurably promoted by higher education. According to McClelland (1976), the main assets of this measure are:

- it makes good theoretical sense in terms of what many people think emotional maturity involves;
- it represents the kind of social and emotional maturity that undergraduate education might well be supposed to influence;
- it is an internally consistent developmental scoring system;

- it indicates changes in relationships to other things that people do which relate to social and emotional maturity; and

- it is highly reliable.

17. Non-Verbal Sensitivity. This test, developed by Rosenthal and his associates at Harvard University (1974), consists of 40 brief voice segments on tape, all of which have been altered to obscure the words. There are two subscales to the test; the RS Subscale, made up of voice segments that are randomly spliced and reassembled, and the CF Subscale, made up of segments which have been electronically filtered so that the words are unintelligible, but the intonation patterns remain. A sample item would consist of a speech segment followed by a question; e.g., "Does the segment represent someone helping a customer or criticizing someone else for being late?" Rosenthal has documented some promising criterion validity for the PONS test. High scorers on this test exhibit the following characteristics:

- they reported warmer, more honest, and more satisfying peer relationships;

- they have been rated by peers and/or by teachers who know them well as being generally more sensitive in interpersonal situations; and

- they were found to be functioning more effectively in the social and intellectual areas of the California Personality Inventory.

18. Moral Reasoning. This test is based on the research in moral development by Lawrence Kohlberg at Harvard (1970). The test consists of a series of paragraphs which describe complex situations in which the actors are forced to choose among several moral courses of action. The task of the applicant is to write a paragraph to justify the alternative that the applicant feels is the best one on moral grounds. The essay answers are scored according to a thematic analysis developed by Kohlberg, and are interpreted according to a schema containing six levels of moral development:

Stage 1: Orientation to obedience and punishment--deference to a superior power or to trouble-avoidance.

Stage 2: Orientation to action that is satisfying to the needs of the self.

Stage 3: Orientation toward approval and to pleasing and helping others.

Stage 4: Authority and social-order-maintaining orientation--"doing duty" and showing respect for authority.

Stage 5: Orientation to duty defined in terms of a contract, general avoidance of violation of the rights of others, and majority will and welfare.

Stage 6: Orientation to high principle or conscience.

The conceptual categories on which the test is based have a high degree of validity as constructs.

By way of summary, Table 1 presents the eighteen measures discussed above in the context of the three general competency domains. A number of these tests have been used in competency-based postsecondary institutions as part of this FIPSE project. All of these measures are being used to some extent in competency-based assessment research.

The following are some of the characteristics and advantages of these measures:

1. These tests require the person being tested to be proactive, not just re-active (i.e., one has to generate responses which can be scored for their appropriateness to real-life situations.) Thus, the test-taker goes beyond recognizing answers, recalling answers, or even generating answers out of context. In the general model, if timing of questions or recommendations is a critical aspect of problem-solving, then this time variable can be programmed into the model as well.

2. The tests are efficient since they can be given to groups as well as to individuals. Their efficiency and economy should substantially reduce the operational costs of current assessment procedures which require vast amounts of time, people, and other resources.

3. These instruments foster equity in the assessment process, since they can be objectively and reliably scored according to the empirically validated coding systems. This is an important advantage since current methods of using juries, panels, or other groups to evaluate are not only inefficient and uneconomical, but are also vulnerable to all the vagaries of subjectivism.

4. The scores can be standardized with reference to criterion groups of which a student is preparing to become a part.

TABLE 1: RELATIONSHIP OF EIGHTEEN COMPETENCY-BASED MEASURES TO THREE GENERAL COMPETENCY DOMAINS

Measure	Cognitive	Effective	Social	Comments
1. Critical thinking	X	x		1,2,3
2. Analysis of Argument	X	x		2,4
3. Concept formation	X			2,3
4. Speed of learning	X	x		1,2,4
5. Learning styles	X	x		2,3
6. Savings Score	X			1,2,4
7. Proactive Case Response	X	x		1,2,4
8. Programmed cases	X		x	2,3
9. General integrative model	X	x		1,4
10. Diagnostic Listening		X	x	2,3
11. Achievement motive		X	x	2,3
12. Socialized power motive		X	x	3
13. Self-definition	x	X	x	1,2,3
14. Stage IV Power		x	X	3
15. Affiliation motive		x	X	2,3
16. Social-emotional maturity			X	1,2,3
17. Non-verbal sensitivity			X	1,2,3
18. Moral Reasoning			X	3

1. Utilized under McBer's FIPSE project
2. Utilized by McBer in other competency-based projects
3. Behavior-referenced and/or construct-validated
4. Pilot instrument; validation in progress

KEY: X=primary relationship
x=secondary relationship

5. Many of these tests tap the competency of "learning how to learn" in a content area. This is one of the most important competencies people can develop because throughout their lives they will be faced with the problem of learning new things in selected areas.

6. These tests are much less threatening and anxiety-producing than traditional tests of recall or recognition, which, because of their properties, only contribute to the fear of failure so prominent in non-traditional students.

7. A number of variations of these tests and the General Model can be developed to add flexibility for administrators, e.g., they lend themselves to video taping, written or oral answers, irdividual or group testing, etc.

8. The majority of these tests have face validity. Educators and students recognize that the skills and abilities being demonstrated are applicable to general life skills.

9. Empirical and construct validation with various occupational and life skills outside of academia means that the competencies required for successful performance beyond the academic program can be established as the target of the learning process.

10. The models and tests can be validated with a variety of non-occupation-specific populations. Some tests and models developed are non-content-specific such that a competent person with little formal education can demonstrate competence as an analytic thinker, information processor, and a pro-active initiator of appropriate solutions. The test format is easily followed and is attractive to those who are test-anxious in traditional test settings.

11. These measures can serve as pedagogical devices as well as assessment instruments, since practice in dealing with the information and component competencies necessary to solve the test problems is a direct way of learning. The instructor and student alike can easily locate and analyze weaknesses and strengths of an individual in exercising component skills. Thus, these measures can serve as diagnostic and guidance tools for supplementary curricular modules.

12. One need not take a particular course or go to a particular college in order to attain competence in the generic skills and abilities measured by these assessment tools.

IV. THE MEANING OF ASSESSMENT MEASURES

Messick (1975) has argued that, until measures have been construct validated, they lack the meaning essential to utilizing them as instruments of general educational theory. McClelland (1973) further argues that, until construct validated measures use relevant real world events among their criterion referents, their value in assessing preparedness for work and life is limited. Educators have often failed to pay attention to construct validity because they "view desired behaviors as ends in themselves with little concern for the processes that produce them or for the causes of the undesired behaviors to be rectified" (Messick, p. 959). In other words, "construct validity is not usually sought for educational tests, because they are typically already considered to be valid on other grounds, namely, on the grounds of content validity" (ibid, p. 959).

In short, educators have traditionally been satisfied with knowing that the content of tests adequately sample a class of situations or subject matter. Messick (1975) argues that content validity does not provide an evidential basis for interpreting the meaning of test scores, and McClelland (1973) argues further that the interpreted meaning of scores that come from construct validation must be strengthened by tying these constructs directly to the world of events outside of academia.

The theoretical distinction between general education and competency-based education is that the latter requires an empirical and causal link between measurement responses and their meaning, as related to real-world life outcomes. Most competency-based programs, however, merely correlate test responses with specific criterion-referenced outcomes (and many do not even do this) without discovering the underlying causes of these responses. Many educators make the mistake of thinking that if a test correlates with a behavioral criterion variable in the world of work or elsewhere outside of the academic world, one can develop competence by "teaching to the test". But this notion confuses correlation with causation, i.e., the fact that tests correlate with observable criteria may only indicate the existence of a causal intervening variable which is really responsible for behavior

and which has not been measured.*

Clearly the mandate for competency-based postsecondary education is to identify skills and abilities that produce (cause) desired outcomes; to develop curricula aimed at the acquisition of these skills and abilities; and to design and validate measures that are sensitive to the acquisition processes and are representative of the criterion outcomes. One should not consider curriculum development apart from assessment issues and neither should be considered in the absence of identified valid performance criteria. Only when these conditions are satisfied does it make sense to teach to the test.

The skills tapped by genuine competency-based tests are largely independent of the content areas in which they are used. For example, the tests for critical thinking, analysis of argument, the problem-solving model, speed of learning, the savings score technique, and other such measures test for generic abilities (competencies) which can be demonstrated in the context of any specific content area. These tests can be adapted to the natural sciences, social sciences, and humanities with equal facility; the content area does not determine the effectiveness of the test. We will always need tests of knowledge, but we also need tests of the way this knowledge is used. The tests we have outlined in Section III satisfy both of these criteria, which represent the essence of competency-based assessment.

Common criticism leveled at the competency-based education movement is that its focus is by definition limited to preparation for specific vocations. A narrow correlational model of competence has fostered this notion, and this concern is legitimate to the extent that criterion validities depend exclusively upon specific job-oriented criterion reference groups. Such validities for liberal arts or general education "are of

---

*For example, vocabulary is correlated with college grades. However, one would not go about improving college grades merely by increasing vocabulary. Doing well in school requires abilities for problem solving, utilizing new information, and other skills not measured by vocabulary tests. Vocabulary is merely a tool, and how it is used depends upon other abilities and characteristics of the individual. One cannot do well in school without a reasonably adequate vocabulary, but having a strong vocabulary will not guarantee success in school without its effective use.

sporadic interpretive utility" at best since they ignore the linking of test behavior to a more general attribute, process, or trait which provides an evidential basis for interpreting the processes underlying test scores. (Messick, 1975)

We strongly endorse this position, but hasten to add that construct validation is itself all too often limited in the types of referents it uses to provide meaning to test scores. Thus, we advocate a validation model that draws from the strengths of construct validation more heavily in the context of real world events or life outcomes than in the context of other constructs alone or "laboratory" behaviors. While Messick (1975) de-emphasizes criterion-referencing, he only does so (1) in terms of using criterion-referents outside of the context of construct validation and (2) perhaps in terms of the type of criterion used as referents. Indeed, all validation is criterion-referenced. The difference in criteria (e.g., "real world" performance, other tests, or observable "laboratory" behavior) determines the extent to which the meaning of the test responses are general or specific and of theoretical or real world significance. A difference between McClelland's (1973) and Messick's point of view is McClelland's emphasis on choosing real world behaviors as opposed to other tests (which typically tap respondent rather than operant behaviors) and laboratory behaviors, as criterion referents. Thus, criterion-referents constituted by a nomological network of life outcomes are consistent with Messick's argument. Espousing such referents differs from Messick's point of view only in terms of emphasizing their selection as criteria for construct validation, not in the validation procedures or concepts themselves. In other words, Messick's notion of construct validation theoretically would include criterion behaviors, but empirically there are differences in emphasis on the types of behaviors to be included. It is for the sake of this difference in emphasis, not theoretical differences, that we have isolated real world events or life outcomes as critical factors in determining the real meaning of tests.

The notion that competency-based education is appropriate for career preparation, but too limited for general education, should have been dispelled by now. The measures developed and used in this project for competency-based education programs have as much applicability for general education goals as for career preparation. Whether one views these measures as appropriate for general education or career preparation depends in all cases on the meaning of the measures, not the measures themselves. And this meaning is determined according to how the validation evidence is marshalled for relating these measures to behaviors, content, constructs, and real world outcomes.

The strength and future of competency-based education rests on its ability to support the rigorous type of research analysis which involves construct validation based heavily upon real world/life outcomes. Until we have identified the critical intervening variables in the causal chain between the educational experience and performance outside of academia, we will be legitimately faulted by critics who view competency-based assessment (and education) as either too narrow in scope or merely "old wine in new bottles."

Epilogue: What's in a Name?

Many people who support the competency-based education and assessment movement do so because it symbolizes a set of values. These values of accountability, relevance, equity, and meritocracy, are the ideological essence of competency-based education. To the extent that people share these values, what is known as the "competency-based" approach to education becomes powerful politically as well as ideologically. There is great danger in this power, and these dangers have already become greatly magnified by those who pay lip service to the new values and ideology, but fail to change their behavior in any significant way from the traditional academic position.

Many educators, who are innovators in educational delivery systems and focus on non-traditional learning outcomes, are ironically hastening the coup de grace of competency-based education more than their "traditional" colleagues. This is because they fail to understand the qualitative differences between assessment procedures or measures that are truly rigorous, reliable, valid, and meaningful--i.e., construct validated and empirically related to real-world outcome--and subjective assessment that is "new" but no more meaningful than traditional techniques. While many educators develop programs under the titles of contract learning, goal-oriented, performance-based learning, programmed learning, experiential learning, and numerous other innovations which espouse the ideology of competency-based education, most of them fail to capture or even recognize the essence of competency-based assessment procedures as construct-validated and criterion referenced.

To the extent that qualitatively superior assessment techniques are the backbone of competency-based education, these innovators have exploited the political power of CBE by appealing to funding sources with ideological rhetoric, and they have diluted the impact of the very changes in educational practice and credibility which they seek. The blame for this dilution of a promising and significant educational movement into a "new" process fad cannot be fairly placed on those whose intentions and practices are good and require support. That is, one cannot blame innovative practitioners who deserve support for their attempts to change and improve the system, for appealing through ideological rhetoric to educational leaders who have financial resources. The "positive reinforcement" for this approach has been too well established by funding agencies.

Indeed, it is axiomatic that the shaping of innovations and their quality is determined almost solely by those who

provide the necessary financial incentives. The educational research and development communities with monetary resources (whether private or public) must be in the forefront of the CBE movement, if significant gains are to be achieved. Support for new assessment technology research and development should not be confused by rhetoric, but in the case of competency-based education it is probably too late. Perhaps, names or titles are too value-laden and politically powerful to be of utility. If so, it is wise that FIPSE is dropping CBE as a funding category.

The true test of relevant actions, however, will be the quality, not merely the direction, of innovation in research, development, and practices that are supported. While FIPSE has distinguished itself by the quality of practices it has funded, the research and development aspects of the process have been neglected. It is our view that assessment research and development must become a priority for the federal government (whether at FIPSE, NIE, or elsewhere) and for private funding agencies if changing practices are to gain necessary credibility and acceptance.

BIBLIOGRAPHY

Atkinson, John W., Motives in Fantasy, Action, and Society, New York, D. Van Nostrand Co., Inc., 1958.

Hodgkinson, Harold L., Evaluation to improve performance, in Vermile, D.W. (ed.), Learner-Centered Reform, Current Issues in Higher Education 1975, San Francisco, Jossey-Bass, 1975.

Kohlberg, Laurence, Stages of moral development as a basis for moral education, in C. Beck and E. Sullivan (eds.), Moral Education, University of Toronto Press, 1970.

Kolb, D., Rubin, I., and McIntyre, J., Organizational Psychology: An Experiential Approach, Englewood Cliffs, N.J., Prentice-Hall, Inc., 1971.

Kolb, David A., On management and the learning process, Working Paper #652, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, 1973.

McClelland, David C. (ed.), The Impact of College on personal characteristics and life outcomes, Cambridge, Mass.: Harvard University, Department of Psychology and Social Relations, 1976 (in press).

_____, Testing for Competence rather than intelligence, American Psychologist, 1973, 28, 1-14.

_____, Power: The Inner Experience, New York, Irvington Press, 1975.

_____, and Winter, David G., Motivating Economic Achievement, New York, The Free Press, 1971.

_____, The Achieving Society, New York, D. Van Nostrand Co., Inc., 1961.

_____, Baldwin, A.L., Bronfenbrenner, U., and Strodbeck, F.L., Talent and Society, New York, D. Van Nostrand Co., Inc., 1958.

Messick, Samuel, The Standard Problem: meaning and values in measurement and evaluation, American Psychologist, 1975, 30, 955-966.

Pottinger, Paul S., Comments and guidelines for research in competency identification, definition and measurement, prepared for the Educational Policy Research Center, Syracuse University Research Corporation, #SURC-TR75-582, 1975.

Rosenthal, Robert, Dane Archer, Judith H. Koivumaki, M. Robin DiMatteo, and Peter Rogers, Assessing Sensitivity to Non-Verbal Communication: The PONS Test, Division 8 Newsletter, American Psychological Association, January, 1974.

Stewart, Abigail, Analysis of argument, in McClelland, D.C., The Impact of College on personal characteristics and life outcomes, Cambridge, Mass.: Harvard University, Department of Psychology and Social Relations, 1976 (in press).

_____, Longitudinal Predictions from Personality to Life Outcomes in College Educated Women, unpublished Ph.D. thesis, Harvard University, 1975.

White, R.W., Motivation reconsidered: The Concept of Competence, Psychological Review, 1961, 66, 297-333.

Winter, David G., The Power Motive, New York, Free Press, 1973.

APPENDIX A

Within the competency-based movement, many innovative approaches to assessment are being developed, many of which borrow from techniques and procedures developed by industrial psychologists. For example:

- portfolios
- journals
- juries
- committees
- life histories
- self-assessments
- supervisor, peer and/or client ratings
- in-basket tests
- work sample tests
- games
- simulations
- projects
- contests
- rehearsed performances

These attempts to break away from the limited traditional measures of verbal ability and scholastic aptitude and achievement have sometimes resulted in elaborate, time-consuming, costly and cumbersome techniques and procedures; and most of these assessment techniques are quite subjective. They are not amenable to standardization for comparability among individuals and institutions.

The major effort underway by ETS (Cooperative Assessment of Experiential Learning--CAEL) to develop new procedures for measuring performance, related to a variety of competencies, is one attempt to break away from traditional measures which are method bound, limited in scope, and of no demonstrable relationship to competent performances outside of academia. CAEL's emphasis on performance measures of learning outcomes is, in itself, a sound approach. However, these new measures suffer from some of the same shortcomings of traditional tests. That is, (1) the techniques tend to be highly subjective and open to broad interpretation; (2) they do not easily lend themselves to standardization across institutions or even among individuals who use them; (3) there is as yet little or no empirical evidence that the performances being measured are any more related to success outside of academia than performances measured by traditional means. Moreover, these new procedures and techniques do not appear to lend themselves to rigorous empirical reality testing, nor to construct validation.

Until a host of measures are developed that are reliable, valid, standardized, construct validated, and rigorously demon-

strated to be directly linked to significant life activities, evaluations and credentials based upon these new performance measures will have little meaning beyond particular institutional settings and will, therefore, not gain wide acceptance.

APPENDIX B

With regard to determining standards of performance, Hodgkinson (1975) stresses the importance of asking good questions about the use and purposes of assessment. Sound judgment and planning are necessary to avoid proceding with evaluative decisions based on ambiguous criteria, standards and/or levels of outcomes. These questions must include: who establishes criteria or standards--an external auditing agency, a faculty member, the institution? What is the reference group with which one will be compared--performers in the real world, students in past years, other students currently being evaluated, one's own past performance, an "ideal" student? What is the proper method of comparison?--norm-referenced tests, criterion-referenced tests, behavioral measures, narratives,(e.g., portfolios, diaries of past experience), unobtrusive measures, etc.? What is the nature of the standard--job performance in the "real world", individual growth and development, ideological ideals of performance, standardized scores? What is the function of the standard?--to select or reject people, to improve performances, to admit students to professional schools or jobs?

If these questions are asked and the answers are concrete, specific and meaningful, a student should know who is judging him, how he will be judged, the nature of these judgments, the objectives related to them, and how well he must perform to meet those objectives.

Two conceptual or technical considerations reported elsewhere (Pottinger, 1975) are also relevant to the issues of establishing appropriate criteria levels of performance.

(a) The Problem of Maximum Levels

Credentials are often restricted to those whose scholastic performance and/or test scores are higher than minimal levels required for work or other social roles. Such occurrences discriminate unfairly against those who are competent to work, for example, but who are selected out of occupational opportunities by those who believe in the simple equation: higher academic achievement means better work or life performance. The tacit assumption that superior abilities in all measured characteristics are necessary or even desirable for performance is highly questionable.*

*A simple motor skill example will demonstrate this point. We know that an automobile driver must grip the steering wheel with enough force to maintain control of the car. But beyond a certain level of pressure, added strength in holding the wheel does not increase overall driving competency. And this is just one of some 3,400 discrete behaviors identified by researchers as making up the task of "driving."

Measures typically used to assess job task performance and performance relating to the mastery of units in a curriculum typically have little bearing on how sub-units interact. For any given job, life task, or individual performance, component skills in one area can compensate for deficiencies in others creating a variety of combinations of individual performance levels which could theoretically "add up to" equivalent overall performance. Thus, minimal levels of performance on individual variables (which compromise overall competence) may have little meaning by themselves. Their interactions with respect to outcomes may have far greater significance.

We are most familiar with this problem in cognitive areas of education. We are often taught language use, verbal reasoning, spatial relationship, reading comprehension, abstract reasoning, and syllogistic analysis (e.g. as measured by Miller Analogies) as discrete units of curricula. Assessment of integrated or general skills such as problem solving often do not take into account the interactive nature of skills in these subcomponent areas. Cognitive measures are used almost exclusively in assessment as if the qualities they measure did not interact, i.e., they are tested separately.

The importance of interactions, while intuitively obvious in the motor skills area, have not been carefully attended to in cognitive and social/emotional areas of assessment. Yet, once individuals have gone through a series of academic life experiences that enhance their competence in dealing with school, work, and other life experiences, the appropriate assessment task becomes that of measuring such integrated and generalized learning outcomes as the ability to cope with new problems, to find appropriate solutions, and to take the correct actions.

Measures which reflect the interdependent nature of cognitive skills essential for satisfactory functioning outside of academia have only begun to be developed.* For example, Klemp's General Integrative Model of Assessment (see pp. 12-13 of text) incorporating a variety of independent techniques, is an approach to summative evaluation of an individual's ability to solve a problem which has as many elements and complexities of real life situations as possible. Such an assessment of individuals has the potential of coming closer to tapping real life competence than can any single test alone.

*A recent example in the noncognitive area by McClelland and Burnham reports the importance of the interaction between levels of motivation and ego-maturity for managerial competence. (Harvard Business Review, Jan.-Feb. 1976)

While it makes sense to require minimal levels of proficiency for many competencies, ability levels over and above necessary cut-off points do not always correlate with overall performance.

For example, in a job analysis, McClelland (1974) found that a minimal level of organizational or clerical competency was necessary for human service workers in the Masschusetts Civil Service system, but high scores on these measures were negatively correlated with superior job performance. Selecting people by rank according to score not only discriminated against those whose scores were adequate (sufficient) though "uncompetitive," but the process failed to select the better job performers as well. This finding and others* suggest that going beyond sufficient levels of competency in awarding credentials can be very dysfunctional for society--not only in terms of equity, but in terms of meritocracy as well.

In many job situations, where cognitive and other competency measures are used to select job applicants, even if job relevance of the characteristics being tested for can be demonstrated (e.g., "verbal ability" in human service workers), level of sufficiency for competent job performance is rarely evaluated or known.

We need more empirical research to establish minimal levels of competence required for quality performance based on how workers in the field perform on various competency measures.

(b) The Problem of Interactions

Researchers have long recognized that the interaction effects of variables are quite often more significant and meaningful than individual variables taken alone. It was stressed earlier, in section III of the text, that competence is not a simple summation of discretely defined skills and abilities. This is readily seen in the example of driving ability. Although one can identify many skills necessary for safe and effective driving--including attitudes, cognitive skills, and emotional factors, as well as perceptual and motor skills--it is intuitively obvious that a simple summation of measurement scores on these discrete task performances would not add up to equivalent driving skills. An individual who is overly competent at some driving skills but woefully inadequate in others would be a poorer driver than someone whose skills were all sufficient, though their summed skill scores would be identical.

*A recent study at Harvard revealed that the past SAT scores of faculty members were negatively correlated with more successful teachers. (McClelland, personal communication.)

The implication for CBE is that one cannot assume that abilities or skills discretely learned will be integrated in work and life functions and consequently that establishment of minimal levels of performance on isolated skills or "sub-competencies" have much meaning in themselves. Therefore, competency research, new assessment procedures, and test instruments must focus more on the interdependence of skills. Basic research as well as empirical analysis of these interactions in various life functions is desperately needed.

Hodgkinson, Harold L., Evaluation to improve performance, in Vermile, D.W., (ed.), Learner-Centered Reform, Current Issues in Higher Education 1975, San Francisco, Jossey-Bass, 1975.

McClelland, David C., Professional Competencies of Human Service Workers, a report to the College for Human Services, New York, 1974.

Pottinger, Paul S., Comments and guidelines for research in competency identification, definition and measurement prepared for the Educational Policy Research Center, Syracuse University Research Corporation, #SURC-TR75-582, 1975.

APPENDIX C

Implications for Research

The implications for research are numerous. The need may be for no less than a new psychology of competence--something on the order of Bloom's and Krathwohl's taxonomies of cognitive and affective dimensions of learning. But the emphasis must be on adult development and learning outcomes with special attention to the interactive nature of psychological variables and how skills and abilities are integrated (as life outside of academia requires). It's a tall order, but a psychology of competence is beginning to emerge.

Most current attempts to define and measure learning outcomes according to what people can do are restricted in scope, lack rigor or poorly correlate with job and life requirements. The current state of the art in assessment calls for more conceptual rigor, more systematic and comprehensive strategies for identifying, operationalizing, and developing measures for new competencies, and more empirical verification of their utility for a variety of life functions.

Until we have a more comprehensive base of empirically identified, clearly defined, and adequately measured competencies, educators will continue to use an existing array of questionable measures based on narrow cognitive outcomes or on a priori value-laden judgments. What is required is a reasonably sophisticated technology capable of uncovering knowledge, skills, abilities, and other characteristics which are necessary and sufficient (as well as "thorough and efficient") for competent performance.

Implications for Change

The heavy emphasis on empirical analysis and verification by researchers should not be taken as a denigration of educators who have strong convictions about what constitues quality education but who are unable to empirically validate these convictions. The intention is not to belittle those who assess student competence on a very subjective basis (that is, "I know competence when I see it"). Clearly, there are many capable individuals in education whose judgements of others are valid and whose evaluation efforts serve students, their institutions, and society well. The plea for more empirical research stems from the belief that such research is critical to the development of quality CBE programs that attempt large-scale change in the way we reach, teach, assess and credential students to assure them more productive and satisfying lives. Moreover, the outcomes of assessment research might well be the "prime mover" in accomplishing the changes desired by those

who view CBE as a major social/educational concept responsive to so many ills inherent in our existing educational system.

CBE will not get far in the endeavor to change this system unless it is able to move beyond what Keeton (1974) has described as a "faddish demand for large scale school change." No matter how strongly such change is supported by those who demand equity and accountability, CBE must provide empirical evidence that it works better than the status quo if it is to become widely accepted. The uphill push against the existing system's reticence to change (as in all systems) will not be sufficiently served by ideological, philosophical or polemical arguments no matter how strongly they side with equity, accountability or other broad social goals. The outcomes must speak for themselves.

Keeton, Morris, Reconnoitering for a CAEL Strategy to Improve the Assessment of the Outcomes of Experiential Learning, in M. Keeton (ed.), CAEL, ETS publication, Princeton, N.J., 1974.